DOCUMENT RESUME

ED 158 749 IR 006 344

AUTHOR Chenhall, Robert G.
TITLE Museum Information Networks. Museum Data Bank Research Report No. 5.
PUB DATE Jan 75
NOTE 21p.; Prepared for the defunct Museum Data Bank Committee

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Electronic Data Processing; *Information Needs; Information Networks; Information Processing; *Information Systems; *Museums; Planning

ABSTRACT

Several museum information processing systems are discussed in relation to their underlying assumptions regarding museum information and networking needs that have resulted in inadequate service to some kinds of museums, particularly historical museums. Rather than any of these systems, a modular approach is proposed that would allow each museum to fulfill its own self-determined information needs in the most flexible and least expensive manner possible. Information processing systems discussed include GRIPHOS (General Retrieval and Information Processor for Humanities-Oriented Studies) and the constraints placed on it by the MCN (Museum Computer Network); SELGEM (Self-Generating Master)--a package of generalized computer programs developed by the Information Systems Division of the Smithsonian Institution primarily for use by museums for general information processing, collection documentation, and for related research-oriented projects; MIRS (Meta Information Retrieval System) and the W.H. Over Museum System--each based upon the same assumption that museum information needs will be served if they have an adequate index to their collections; and GIPSY (Generalized Information Processing System) which assumes that rapid retrieval and display of specific records is the most important need to be satisfied. (Author/MBR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

- ED158749

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MUSEUM DATA BANK RESEARCH REPORT

Number 5

Museum Information Networks

by
Robert G. Chenhall

January 1975

This report, or any portion thereof, may be reproduced for any non-commercial purpose provided appropriate credit is given to the author and the Museum Data Bank Committee as publisher.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. G. Chenhall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Museum Data Bank Committee
Robert G. Chenhall, Chairman

Margaret Wocdbury Strong Museum
700 Allen Creek Road
Rochester, N.Y.

IR006 344

The title of this paper is somewhat misleading, for there are no museum information networks in existence today, if the term network is used in the usual sense to mean an interconnected configuration of computers and computer files that allow a user to obtain data from any of the files if he is in a position to initiate a request from any of the computers.

Whether or not a museum information network is a desirable objective depends upon the circumstances and the needs in different types of museums. The position expressed in this paper is that the needs of museums generally are best served if the three subjects of museum information, information processing systems, and network processing systems are considered separately. Further, it is the opinion of the author that the information needs of some kinds of museums (particularly historical museums) have not been adequately served because certain assumptions concerning museum information and networking have been incorporated into the computer systems that are presently being used for processing museum information. In order to explicate this position the assumptions apparent in some of the systems will be discussed. Following this, a suggestion will be made for a modular approach that would allow each museum to fulfill its own self-determined information needs in the most flexible and least expensive manner possible.

MUSEUM INFORMATION PROCESSING SYSTEMS

Museum information processing systems are in all cases examples of what are called, in more generic terms, information systems or information storage and retrieval systems. There are areas of museum management that require other kinds of computer capabilities (e.g., payrolls, counter sales, inventories, and so on). However, the problems of collections management and curatorial research loom so large in most museum environments that the term museum information has come to be almost synonymous with information about museum collections, and this type of data can most efficiently be handled with computerized information systems.

The systems known as GRIPHOS and SELGEM are undoubtedly used by more museums than any of the others. The major discussion here will be concerned with these, although some of the characteristics distinctive of others will be considered more briefly. All of these systems have certain strengths and weaknessess when considered for a particular application. None of them are sufficiently generalized to perfectly serve all of the needs of all museums everywhere.

GRIPHOS and the MCN

The term GRIPHOS is an acronym for the computer package entitled General Retrieval and Information Processor for Humanities-Oriented Studies.[1] The program package was developed by Dr. Jack Heller, a

---

1. See Chenhall 1975 for more information concerning GRIPHOS and most of the other systems discussed in this paper.

computer scientist at the State University of New York at Stony Brook, largely under contracts with the United Nations. The use of GRIPHOS in museums has been through the agency of the Museum Computer Network, Inc., and many persons mistakenly assume that GRIPHOS and MCN are one and the same thing. However, it must be pointed out that most of the constraints placed upon the museum use of GRIPHOS are not the result of assumptions about museum data that are built into GRIPHOS. Rather they are limitations imposed by the MCN application of GRIPHOS.

The MCN began in 1967 as an informal consortium of sixteen museums--mainly art museums in New York City. David Vance, now president of the Museum Computer Network, Inc., stated in an early article (Vance 1970: 204) that the informal seminar directing the work of this consortium through the spring and summer of 1967, "...was guided by three main considerations: (1) the capabilities of Prof. Heller's GRIPHOS macro language ..., (2) the nature of existing museum records and (3) the information needs of museums and their scholarly public". Some of the present (1974) characteristics of GRIPHOS will be discussed below. For the moment let us consider the assumptions which developed out of considerations (2) and (3).

In his discussion of existing records Vance points out that museum records are anything but standard. However, he goes on to say (Vance 1970: 205; 206), "Underlying this apparent diversity is a remarkable uniformity of factual content. The possibility of an organized data bank derives from the presence of comparable data, implicit or explicit, in nearly all records.... It has already been demonstrated that enough natural order exists in randomly collected records to allow their incorporation into a systematic if somewhat irregular structure."

As to the information needs of museums, Vance (1970: 205) states that: "The data bank's main objectives are to locate objects and to bring information under control." He goes on to point out that an object is effectively lost when an interested scholar cannot find it easily and that the control of recorded information in the past has been possible only on a small scale because of the complexity and cumbersomeness of crossfiling massive files.

The underlying assumptions concerning museum information which are stated above and in more recent writings of Vance (e.g., 1973; 1974) are:

1. The eventual objective of the MCN (and, presumably, of all museum data bank activity) is to create some form of network that would make the data on objects in a museum available to anyone who is a part of that network.

2. There is sufficient 'natural order' or 'linguistic structure' (both concepts are suggested by Vance in different places) in museum records so that acceptable rules or conventions can be stated for the recording of any object, person, document, place, event, or concept (Vance 1970: 206; 207; 1974: 3).

3. In order to make it feasible to eventually create a true museum information network it is essential that all members of the MCN record information in their own computer files according to the rules or conventions of the MCN. For example, the MCN has defined a series of tag numbers to be used with the recording of objects, persons, documents, and places. "Since each record describes something belonging to <u>one</u> of these types, tags for more than one type may not be mixed in the same record

(Vance 1974: 4)."

4. In order for a museum information network to work it is essential that all members have exactly the same program package--i.e., it is assumed that each MCN member must be capable of communicating data directly to (or receiving data directly from) every other member.

It is, of course, possible to create a museum information network on the basis of these assumptions. Each of them, though, places certain constraints upon the individual museum that is a member of the network (i.e., that has access to the GRIPHOS program package), and unless they are a necessary part of networking it appears doubtful that they will stand the test of time.

Some of the questions that may be raised about each of these assumptions are:

1. The assumption that a physical network of some kind is the eventual goal may be accepted without question by art museums and others concerned with objects that are widely dispersed and frequently must be gathered from a variety of sources for exhibition or other purposes. Most historical museums, however, are more interested in information processing systems that will enable them to bring their own collections under control than they are in any kind of information network. There is no objection to the idea of interfacing data banks eventually but not if networking makes it more difficult to satisfy the information needs of the individual museum now.

2. The assumption that there is either a natural order or linguistic structure implicit in any verbal description of an object, person, document, place, event, or concept is a philosophical question which could be argued at length. It is the opinion of the present author that a large portion of any apparent order or structure is imposed upon the entities by the immediate purposes and the theoretical perspectives of the individual responsible for the data file. It has been our experience in dealing with archaeological data files that the purposes or objectives of every data bank installation (and, in fact, every data bank file) must be stated clearly, accurately, and precisely, and the data structures must be developed with those purposes in mind, or the file, once created, will not be used (see Scholtz and Chenhall n.d.). To some extent the structuring of museum information is arbitrary and a structure imposed by the MCN is as good as any other such structure. However, there are many indications that: (a) no predetermined set of tag definitions can be completely adequate to serve the information needs of all museums; and (b) whether an object is one or another of the MCN entities (i.e., object, person, document, place, event, or concept) in some cases will depend upon the perspective of the individual recording or retrieving the data. As just one example of this we might ask whether a photograph is a document or an object? The answer, of course, is that it may be either one. This cannot be determined until the time of retrieval and why should it make any difference anyway?

3. The third and fourth assumptions of the MCN--that a network operation is feasible only if all members conform to the same data entry rules and have exactly the same program package--are valid only if one assumes that a single program package will fulfill all of the functional requirements of both a museum information processing system and a museum information network system. It is equally feasible and ultimately more productive, I believe, to envision a museum information network as a functionally separate operation, capable of accepting input from and transmitting data to any number of different museum information processing systems (see Greenberger, et al. 1974). It then becomes necessary for the network processing system to interpret and convert as necessary from the data recording conventions of one processing system to those of another system. Conversion, though, logically can be a subsidiary part of network processing. It is not something that must be avoided by having everyone use a single, monolithic information processing system.

When the assumptions regarding the nature of museum information that are implicit or explicit in the MCN are separated from the assumptions implicit in GRIPHOS, it is possible to identify the latter quite clearly.

1. One important assumption about the nature of museum information that is implicit in GRIPHOS is that word strings are more important than the statistical processing of numbers. As a direct outgrowth of this assumption data is entered into the system in an unstructured, free-form mode that can be readily sorted from left to right in the manner of an alpha-numeric index. This, however, is a different mode of data storage from the right justified, fixed form mode that is necessary for statistical data processing. One of the later developments of GRIPHOS is the capacity to convert numeric data from one mode to the other in order to interface with statistical program packages. However, statistical data processing will always be cumbersome and inefficient with GRIPHOS simply because the system was created with the primary emphasis on ease of data string analysis and the flexible formatting of output reports.

2. A second assumption of GRIPHOS concerning the nature of museum information needs is that most of these needs will be met by searching, sorting, and printing (in an attractive format) data derived from the initial input file. It has always been possible to create inverted indexes with the GRIPHOS system but except for the frequency dictionary report (an important facet of input editing)these have not been widely used by museums because they were not coordinated with and made an integral part of the other processing functions--i.e., data entry, data corrections, etc. In a historical museum an accurate, up-to-date, and readily accessible index is often a more important source of information than the complete records in unsorted sequence. Therefore, the ability to correct an inverted index concurrently with any modification of the indexed records (a new innovation which we understand has now been implemented) is a definite improvement in the GRIPHOS capability.

3. The idea of linking or 'threading' from one file of records to another automatically (as for example, in being able to access an artists' biography file by means of a tag for the artist's name in an art object file) has been an important concept of GRIPHOS from the very beginning and it has many applications in museum information processing which so far have been only partially explored.

4. The necessity for hierarchical data structures in order to avoid excessive redundancy in the initial recording of musuem information (as well as the excessive file space required to store the redundant information) has also been recognized as an assumption in GRIPHOS from the beginning. The methods used to link sub-files up or down through a hierarchical GRIPHOS structure are much the same as the methods used to link unrelated records (see Heller 1974).

SELGEM

SELGEM, an acronym for SELf-GEnerating Master, is a package of generalized computer programs developed by the Information Systems Division of the Smithsonian Institution primarily for use by museums for general information processing, collection documentation, and for related research-oriented projects. It was created by Mr. Reginald A. Creighton, Manager-Information Retrieval and Indexing Group, and Mr. James J. Crockett, Manager-Software Systems and Program Maintenance, out of an earlier system called SIIR (Smithsonian Institution Information Retrieval). At an early stage (c. 1970-71), SELGEM received substantial support from the National Museum of Natural History. Many of the programs were written primarily to meet the needs of natural historians, although they are now used for other purposes as well.

As contrasted with GRIPHOS and MCN, SELGEM is a "single level" system. That is, it does contain certain assumptions regarding the nature of museum information, but these assumptions are all implicit in the program package itself; there are virtually no constraints placed upon individual users by the loosely structured user organization. When a user acquires the SELGEM system from the Smithsonian Information Systems Division, he normally receives a source program package written in COBOL. The user must assemble the program and, if necessary, rewrite the COBOL routines so that they will function on his particular brand and model of computer. However, he is also free to modify any of the programs as he wishes and/or to write programs of his own in order to adapt the system to his particular needs.

Some of the assumptions that are implicit in the programming philosophy of SELGEM are as follows:

1. Contrary to the assumption of the MCN, SELGEM does not assume that the ultimate objective of museum data processing will be to create integrated networks. SELGEM is an information processing system only. No attempt is made to control either the information that is recorded using the system or refinements to individual programs that may be made by the users of the system.

2. With SELGEM there is an assumption that statistical data processing (i.e., counting and mathematical manipulation of the data) will occur often enough so that it should be possible to enter data in a format that is useful either for numerical processing or word string processing. Input records thus are structured in a fixed format corresponding to eighty column cards placed end to end on magnetic tapes. This makes it possible to retrieve records that contain numerical data and to process them directly as input into statistical programs without

any of the data conversion problems that are required, for example, to do statistical processing with GRIPHOS records. At the same time each SELGEM record contains space from columns 16 to 79 in which data can be structured in any manner desired and, if more space is needed, this can be carried forward or continued from record to record so that it is possible to record lengthy strings of alpha-numeric data as well. The simplified data structure gives the user the advantages of both statistical manipulation of numeric data and string manipulation of alpha-numeric data.

3. The developers of SELGEM have assumed that the system should be as understandable as possible. The evidence for this is: (a) The fact that the COBOL programs are available in source language form so that the user who has even a moderate amount of programming knowledge can understand not just the result of a particular program but also the logic of how each step is accomplished; and (b) the fact that records are structured as images of eighty column punched cards. There does not appear to be any mystery about data structured in some manner other than as a punched card, but it is a fact of life that the moderately informed user better understands and identifies with data that is structured to resemble a card.

4. SELGEM assumes that the total system of computer programs should be capable of being modified as necessary to meet the information needs of individual institutions. This does not necessarily mean the rewriting of individual programs, although that is permissible. It also provides for the ready interface of Smithsonian-written programs with individualized programs that may be created by the user. This is the complete opposite of the MCN assumption regarding the need for networking and, thus, the need for everyone having the same program package. It can also be considered as a step toward the modular approach suggested in the final section of this paper.

5. It was pointed out above that GRIPHOS assumes that most data needs would be satisfied through searching, sorting, and printing data derived from the initial input file--i.e., that inverted indexing was of secondary importance. SELGEM makes the same assumption. Data frequency reports are a form of indexing that is regularly used during input processing; and it is possible to prepare key word indexes from data that was previously entered. However, key word indexes have proven to be useful only where there is relatively little change in the file. Additions, corrections, and deletions that are made in a data file cannot be simultaneously added, corrected, changed or deleted from an index file. If it is important enough to have an index for a file, that index must be reconstructed each time the file is changed. With a large file this can be a major problem.

6. The developers of SELGEM have recognized the importance of hierarchical data processing as a means of reducing redundancy of data input and storage. However, the hierarchical structuring of records in SELGEM comes about through an ascending numerical relationship among the serial numbers, category codes, and line numbers that are assigned to individual records--i.e., through the structuring of the data itself rather than through any kind of programming capability. No provision is made for

automatically linking one record to another (whether or not the records are hierarchically related) except through the processes of sorting and merging. In a historical museum, where there are numerous instances of objects which in themselves are important but also are parts of sets or otherwise are related to other objects, the ability to link records through program threading would be a desirable attribute.

The three most distinguishing characteristics of SELGEM are its ready availability, its understandable simplicity, and the fact that it is relatively inexpensive to operate. Because of these characteristics SELGEM is a good system to use as a point of beginning, to bring a collection under control. As files become progressively larger, though (as inevitably they must) and as museums become more sophisticated in their information demands, the need for adequate inverted indexing and the need for hierarchical data organization will become increasingly critical. A separate indexing package (or "module"; see pp. 11-13) can probably be used as a companion to SELGEM to meet the first of these needs; it is my opinion that neither of them can be adequately met within SELGEM because of the way the records and files are structured.

MIRS and the W. H. Over Museum System

MIRS (Meta Information Retrieval System) was designed by Mr. Robert Shapiro, president of Meta Information Applications, Inc. of Wellfleet, Massachusetts to operate on a PDP-10 computer. It is a versatile general purpose retrieval system that has been widely used in business. The only museum application of MIRS at present is at the Children's Museum in Boston.

The unnamed system used by the W. H. Over Museum at the University of South Dakota is an example of what can be done in the preparation of a computerized catalog by means of special programming for a small scale computer. It operates on an IBM Model 1620-1 with a core size of 40K and the files are stored on IBM Model 1311 magnetic disks. The programs were written by a student at the university, Mr. Gerald Norby.

The computer configurations required to operate these two systems are about as different as any two computer systems could possibly be. However, both of the projects are based upon the same assumption concerning the nature of museum information: namely, that a museum's information needs will be served if it has an adequate index to the collection. Thus a computer record about a specimen can consist of only two parts: (1) a free-form textual description of the object, and (2) a list of key word descriptors that might be used to find that record. The structuring of the record is essentially free-form, which means that a record can be of almost any length and that there are no predefined fields within the record, either created by the use of tags (as in SELGEM and GRIPHOS) or through positional structuring of fields within the record. In the W. H. Over Museum system the list of key index words always precedes the descriptive text and is separated from it by a single asterisk. Two records in a computer file are separated

by a double asterisk. Records are stored on the disk in random order, and when a user of the system wishes to retrieve all the records on objects which contain a specific word as one of the descriptor terms, the system searches each record sequentially to determine the presence or absence of that term in the list of descriptors. Each time there is a 'hit' the record is set aside and at the end of the operation all of the records which contain the desired word are listed on the printer in the computer center.

With the MIRS system an inverted index is updated each time a new record is entered on the remote teletype terminal. When a search is instituted, also on the remote terminal, for all records that contain some particular word or combination of words the index is searched first to determine the number of records with the required terms and the storage location of those records. Upon receiving additional instructions the computer displays on the remote terminal the full paragraph form text of all of the records retrieved.

The developers of SELGEM and (until recently) GRIPHOS have played down the importance of indexing in the total scheme of museum information needs, and have emphasized instead the searching of the full file for either all data or particualr strings of data that are stored in conjunction with 'tags' or coded field designators. In contrast to this MIRS and the W. H. Over Museum system assume that a single index to the records is the only museum information that is needed. The viability of any of these assumptions, of course, can only be determined by the relative numbers of institutions that are willing to adopt the systems based upon each set of assumptions.

GIPSY

GIPSY, an acronym for Generalized Information Processing SYstem, is a program package developed in 1968 under the direction of Dr. James W. Sweeney then director of the Merrick Computer Center at the University of Oklahoma. The first use of GIPSY in museums was a project in which a centralized ethnographic data bank was created for all of the American Indian artifacts in Oklahoma museums. The University of Missouri also used GIPSY to develop its own ethnographic data bank. These projects have since become more or less inactive. However, the Oklahoma Biological Survey is now using the system for the development of a data bank of biological specimens. Emphasis of the project has been to collect environmental data and longitude/latitude coordinates so as to facilitate retrieval for environmental impact statements.

The assumption about museum information which is implicit in the structuring of GIPSY is that rapid retrieval and display of specific records is the most important need to be satisfied. Thus, the most effort and the most sophisticated programming was devoted to providing a mechanism by which the user could sit at a remote terminal and search the computer file in an interactive, conversational mode, using Boolean logic

where necessary, and examine the results in a terminal display immediately.

GIPSY has not been widely accepted by museum people (although it is actively being used in other fields), and the ethnographic data files that were created using the system have to all intents and purposes fallen into disuse. The reasons for this, however, do not appear to be the result of any erroneous assumptions concerning the nature of museum data. What Dr. Sweeney initiated originally was a soundly conceived system, but: (1) the persons responsible for the museum applications moved on to other activities soon after the original files were created so that for a period of several years there was no mechanism of communication between interested museum people and the computer analysts who could have made this system more responsive, (2) at about the same time Dr. Sweeney also moved on to other interests and for several years there was a void in the creative leadership that is required to continuously update and improve a system, and (3) perhaps because of the above the University of Oklahoma also lost interest for several years and made no attempt to encourage the use of the system by museums. A further problem was the fact that an effective system of inverted indexing was never integrated into GIPSY as a regular part of data entry and data correction so that with a file of no more than 20 to 30,000 records the response time for the first search of the file (of necessity the complete file) became inordinately slow.

GIPSY could be developed into an effective museum information processing system--comparable in scope to SELGEM and GRIPHOS. However, it has suffered from neglect now for several years and if it is ever accepted again someone who is both respected in the museum field and knowledgeable about computers will have to spend several years rebuilding the momentum that has been lost.

Other Systems

In England the Information Retrieval Group of the Museums Association (IRGMA) has been encouraging and to some extent supporting the development of a generalized museum communication format by John Cutbill at Cambridge University. The format has been implemented with an information system called CGDS (Cambridge Geological Data System). However, Cutbill has stated privately that CDGS will not do everything that a museum information processing system should do and he is presently working on a new version that will accomplish the desired objectives.

In his written publications, Cutbill (1974: 3-5) is critical of the assumptions of the Museum Computer Network data structure--particularly the MCN assumption that all data can be expressed by (1) mutually exclusive data category definitions (i.e., tags) and (2) the content of data recorded using those tags. The data structure suggested by Cutbill as an alternative is considerably more flexible than the MCN data structure because, in theory at least, it allows one to link related data

elements within a single record (e.g., dimension value and unit of measurement) and it allows one to nest subsidiary or qualifying data categories in the form of hierarchically linked records. It does not appear, though, that Cutbill recommends anything that cannot be done within the present GRIPHOS program package capabilities. Rather his criticism is directed toward the single level form of data structure contained in the MCN use of GRIPHOS.

It is difficult to evaluate the museum communication format which has been proposed by Cutbill. There is no question but what this has worked well with geological data using the CDGS. Also: (1) It is becoming increasingly apparent that museums will have to give a great deal more attention in the future than they have in the past to alternate ways of structuring data, and (2) Cutbill's format does permit one to handle some of the problems that are raised in the MCN use of GRIPHOS. However, it appears as though the museum communication format makes the same assumption concerning museum information that is made by the MCN data format: namely, that a network operation is feasible only if all members conform to the same data entry rules and have exactly the same program package. These assumptions are valid only if one assumes that a single program package will fulfill all of the functional requirements of both a museum information processing system and a museum information network system.

The organizational support which Cutbill has received from IRGMA and the financial support he is now receiving from the Office for Scientific and Technical Information (OSTI) should provide him with the resources necessary to produce, within the next two to three years, the first of a "second generation" of museum information processing systems. Hopefully, the weaknessess in the existing systems will be overcome in this effort.

TAXIR (an acronym for TAXonomic Information Retrieval) and the version of TAXIR currently in use at the University of Colorado known as STIRS (Set Theoretic Information Retrieval System) were developed as a computerezed system for systematic biology. They are based upon the assumption that strings of verbal data, such as those normally processed by the systems described above, cannot be handled rapidly enough to satisfy museum information needs. Because of this assumption the data input to TAXIR and STIRS is entered into a master file in the form of a highly coded matrix using the mathematical principles associated with set theory. This master file matrix, when resident in the computer memory, contains a record of all data on each item across the horizontal axis and fields or descriptor states along the vertical axis. The contents of the matrix are represented by bit strings (ones and zeroes). All processing is done (very efficiently) with the complete master file resident in memory. At the conclusion of a processing cycle the complete file is written out onto a magnetic tape for storage.

It is difficult to assess the probable future of TAXIR. It would appear as though the system requires the user to give more attention to the detailed, set theoretic file structure than most museum people will be willing to give to it. In fact, in areas other than science museums the mental set of most curators might preclude them from ever understanding how TAXIR works. The system is being used in a number of fields outside of museums, however, and it remains to be seen whether it will be widely accepted for museum information processing purposes.

GIS (Generalized Information System) is a generalized computer program product of the IBM Corporation which was designed for building and maintaining large collections of information and querying and retrieving the stored information. GIS became known to museum people largely because it was adopted by the now inactive Flora North America Program as a basis for maintaining their centralized files. There are many different options associated with GIS and it is virtually data independent. There are, of course, no assumptions in GIS concerning the nature of museum information needs and the system was adopted by the Flora North America Program primarily because of its exceptionally fine capability in the processing of hierarchical files.

The Museum Data Bank Committee now has files on sixteen different information processing systems, in addition to those that are used only at a single institution. The systems known as SOFIA (Gardin and LeMaitre 1974) and ADAM II (Gaines 1974) are examples of information processing systems developed, in France and Arizona respectively, for the specific purpose of processing archaeological data. The assumptions in these two systems, thus, include elements of archaeological theory as well as assumptions about museum information. Other systems are being developed in the Netherlands, Sweden, West Germany, and as far away as South Africa. The lengthy discussion above, however, is sufficient to illustrate the variety of approaches that have been taken.

A MODULAR APPROACH TO MUSEUM INFORMATION NETWORKS

The evolution of computers and computer systems generally has been characterized (e.g., by Booth, 1973: 4-13) as a gradual shift from applications in which integrated packages were individually designed for each new installation toward applications in which the analyst puts together special purpose modules that are designed to perform specific functions. These modules may be different hardware units or they may consist of various kinds of software systems. The important thing is that they are created or manufactured to perform specific functions as part of larger systems and they are capable of being interfaced with a variety of similar special purpose modules in any given installation. Hardware modules may be such things as the disk pack units of one manufacturer which are combined, perhaps, with a core memory produced by another manufacturer, and possibly an input system of a third manufacturer. Examples of software modules are: generalized languages such as assemblers, compilers, and data management languages;

editors, loaders, and general purpose I/O routines; file management packages; task management packages; resource management packages; statistical control routines; and supervisory control routines. All of these are functional modules which in theory and increasingly in practice, can be taken off the shelf and put together into a total configuration that serves the needs of a particular installation in the best--and cheapest--manner possible.

This idea of functional modularity has only recently been expressed as a principle of network design and processing system design (in a pair of companion volumes by Booth, 1973, and Becker, 1973). However, the fact of increasing interchangeability of computer operating systems and program modules has been an observable phenomenon ever since 1951 when the first commercial computer installation, the UNIVAC I, began operation at the Census Bureau.

Most of the museum information processing systems in use today were designed in the late 1960's. By then anyone acquiring a computer could assume that he would obtain, along with the hardware itself, many of the software modules enumerated above. These software modules often were designed (as they still are) to function only on the equipment of a particular manufacturer and at times only on certain models of that equipment. This is a long step from the situation in 1951, when all of the programs to run the machinery had to be laboriously written out in binary coded machine language, but it means that most of the systems described in the early part of this paper had to be designed as packages of programs that would accomplish, on a particular computer, the entire spectrum of information processing from data input, through editing, to final printed or terminal output.

All of the program packages used by museums today are portable in the sense that they can be moved to another installation where the computer hardware and operating software are approximately the same; and some of the program packages used by museums are portable from one brand of equipment to another by virtue of having been written in a more nearly standardized language such as FORTRAN, COBOL or ALGOL (although the latter type of portability is more apparent than real due to the lack of standardization among the assembler languages used on different computers; months of reprogramming are often required in order to adapt a program to a different brand of equipment). None of the program packages used by museums today were designed as special purpose modules that would function as an adjunct to or in conjunction with some other program package.

The central thesis of this paper is that a modular or functional approach to the subject of museum information networks ultimately will serve to meet the information needs of museums in the most flexible and least expensive manner possible. In order for such an approach to be feasible the developers of the different systems will have to agree on certain minimal conventions. This, however, is not an impossible dream. A meta-language already exists to convert data between GRIPHOS, SELGEM and GIPSY.

A modular or functional approach to the subject of museum information networks has many implications for the potential user. Most important is the fact that in any given institution the three general areas of museum information, information processing systems and network processing systems could be considered separately, and modules could be selected from a number of systems if necessary in order to fulfill precisely defined needs. Let us look at each of these three subjects separately and with a few examples attempt to show how this might work.

Museum Information

The most difficult and time consuming task involved in the creation of a viable museum information system does not involve computers at all. It consists of four activities which I believe should be performed in a given order and, as nearly as possible, without regard for the techniques that will be used to accumulate, record, control and retrieve the data. These activities are:

1. Definition of purposes. I have expressed the opinion (p. 3 above) that the purposes or objectives of every data bank installation (and, in fact, every data bank file) should be stated clearly, accurately and precisely or the file, once created, will not be used. Numerous examples, good and bad, could be given to illustrate the need for clear, precise and objectively defensible statements of purpose. Generalized expressions such as, "It will enhance the value and potential use of collection-related data," or "It will provide a number of collection-management benefits," are not adequate.

2. Delineation of needed information. In most museums a great deal of information is accumulated about each object. It is assumed that all of this information will eventually be needed and, therefore, that it should be stored in any computer file. Seldom does anyone take the time to explicitly delineate the information that will be necessary in order to accomplish the defined purposes. Where this has been done greatly simplified and often relatively inexpensive information systems have evolved.

3. Delineation of data structures. Once the needed information is clarified, it is possible to formulate data structures that will provide this information in the most economical manner possible. The implications of this are explored in depth in Scholtz's (1974) article on "Data structure and computerized museum catalogs."

4. Creation of thesauruses or authority files. If information is to be recorded in a consistent manner, it is essential that data categories which will be used for indexing and retrieval be recorded initially using both recording conventions and terminology that is consistent from one record to another. Internally consistent thesauruses are essential to the achievement of this objective.

Very little work has been done on the exploration of the content of museum information apart from the methods that are used to control and process that information. In a modular system such as that proposed each institution would be forced to consider its own needs carefully rather than having those needs either determined by or limited by assumptions that are built into information processing systems.

Information Processing Systems

It is only after information needs have been clearly defined that a system can be created to serve those needs adequately. I use the word "created" here to further imply the selection of system modules an as alternative to adopting a unified, monolithic information processing system. Two examples will clarify the distinction.

The first example is a natural history museum in which the defined data purposes are (1) to control the collections within the museum, and (2) to be able to prepare environmental impact statements as easily as possible. The information necessary to accomplish the first objective can be provided by computerized listings of key categories of data about each item. To accomplish the second objective it is necessary that the location where each specimen was collected be precisely recorded within a system of geographic notation that is mathematically consistent, and that there be some way of easily retrieving the records on all specimens that were found within any specified geographic boundaries.

The functions of data entry, input editing, retrieving, sorting and listing specimen-related data can be accomplished using any one of several different processing systems that are now available. If a decision is made, however, to record the location where the specimen was collected using the UTM (Universal Transverse Mercator) projection, the one missing ingredient would be a computer program that would accept as input the geographical boundaries of an area and from this retrieve all specimens found within that area. If a program module to accomplish this had been previously written the user should be able to select that module and interface it with the rest of his system regardless of the basic processing system (i.e., SELGEM, GRIPHOS, etc.) he is using for the routine functions of data entry and so on.

A more complex example involves a relatively large history museum, one which has, perhaps, 500,000 artifacts in its collections. Here the defined purposes are entirely a matter of internal control: the ability to locate different classes of objects, or specific objects, for research or exhibit purposes; the ability to readily locate the records concerning a specific object that is in hand; the ability to provide lists of objects in given storage or exhibit locations; the ability to summarize appraisal data for insurance purposes; etc. To accomplish these objectives requires a great deal of work in the areas of delineating needed information, data structures and authority files. In addition, it requires an information processing system which, among other things, includes a readily accessible and dependably current index to certain categories of data and the ability to process data hierarchically when necessary. Indexing is a complex operation. A program module to accomplish it effectively

might function as a separate sub system that produces nothing more elaborate than a terminal display and pointers to indicate (internally ?) the indexed records. Once developed, such a module should be capable of being interfaced with any of the the extant information processing systems. I realize, of course, that it would be necessary to reprogram this module in order to make it work on different brands of equipment. However, the module should be designed so as to work equally well with SELGEM, GRIPHOS, GIPSY and as many of the other systems as possible. It should not be designed to be a proprietary part of just one of these systems.

Network Processing Systems

Very little has been said here concerning the subject that is the title of this paper: Museum information networks. There are two reasons for this:

1. The need for a physically connected network of computerized museum data files has not been demonstrated. There may be instances where such a need does in fact exist, but careful studies of the information needs in different types of museums--the type of studies suggested above under "Museum Information"--will have to be conducted by a number of museum organizations in order to demonstrate this need.

2. A model for the modular approach to network processing systems is already available. This is discussed at length elsewhere (Chenhall 1975: 246-250; see also Booth 1973 and Greenberger 1973, 1974). The best example is the ARPANET, which provides for the interchange of information between defense contractors who individually use a wide variety of computer hardware and computer software systems. The programs necessary to interface between these systems are simply additional modules in the overall network operation.

The longer I work with museum data banks the more convinced I become that we are addressing the wrong problems first when we start with an assumed need for either a network or a regional or nationally centralized data file. Networks may come in time but the subject is not of first order importance to the museum community generally. It is true that with some types of museum collections it would occasionally be desirable to be able to integrate some of the information from a number of different individual catalogs. The art museums have been mentioned as having a genuine need for this kind of integrative ability; and those who are associated in any way with environmental impact statements will readily recognize the desirability of being able to integrate extant information about given geographic areas, wherever it may be filed. Networking may or may not be the best means of accomplishing this kind of data integration (There are some alternatives). However, the internal needs of individual museums represent a far greater challenge at present than that posed by the need to integrate limited amounts of information from a number of museums.

For those situations where museum information networks are ultimately the best answer the state of the computing art today seems to clearly

dictate that network processing should be separated from information processing. Becker (1973) and Greenberger (1973, 1974) both define, in slightly different ways, the functions that are included in network processing. If even a small portion of these functions are imposed upon the processing of museum information within individual institutions, the cost effectiveness of computers will not be sufficient to be appealing to very many museums.

CONCLUSIONS

Computer systems are, in some respects, like automobiles, houses, works of art or any of the other artifacts of man: at the time they are created it is difficult (especially for the creators) to think that the newness, the usefulness, is not eternal. We know that time takes its toll and that what is useful today will no longer be adequate tomorrow. Still, we would like to think that our particular creations will be around long after we are gone.

In the field of computer science major conceptual developments have required the scrapping of all that went before. The idea of a second or third "generation" of computers carried with it not just more advanced hardware, but completely new programming concepts and the necessity for conversion or reentry of data. The field has recognized, however, that this cannot be carried on indefinitely. Economics has dictated that further changes will not require massive reprogramming or data reentry each time a change is made. The concept of modularity has slowly evolved as a means of anticipating and dealing with change (either change to accomplish a greater internal efficiency or change as a modification of data needs) without creating an entire new structure.

The information needs among museums are not the same as they were when computer systems such as SELGEM and GRIPHOS were designed (I will not argue whether the needs may not be the same; at least they are defined somewhat differently today). The data needs in museums, as defined today, could progably be accommodated by expansion or modification of existing systems. The important thing, however, is that we must anticipate continuing expansion of these needs and we must provide a way of accommodating these changes as inexpensively as possible. I have mentioned several areas where I consider expanded needs to be evident today--e.g., inverted indexing, a uniform system of geographic data retrieval. I seriously believe it would be a disservice to the museum profession if computer programs were written to fill these needs within just one of the present information processing systems.

An indexing system, for example, might be designed (perhaps with slight variations to make it usable on all of the major computer hardware) to work on the front end of any data entry or data correction programs. An indexing package requires a minimum hardware configuration that includes disks or slow-core storage (i.e., it cannot function with tape storage). However, it is entirely conceivable to pass data through a FORTRAN indexing program, for example, and then immediately into a SELGEM or GRIPHOS data entry program. There would be some problems with this, of course, and it might be

necessary to convene special working sessions in order to be sure that such a module would in fact be compatible with the major existing systems. However, it seems to me that economic necessity is going to increasingly demand this kind of cooperative, modular approach just as in the early part of the nineteenth century it forced the railroads to finally agree upon a standard track width of five feet eight and a half inches.

REFERENCES CITED

Becker, Hal B.

1973 Functional Analysis of Information Networks. John Wiley & Sons.

Booth, Grayce M.

1973 Functional Analysis of Information Processing. John Wiley & Sons.

Chenhall, Robert G.

1975 Museum Catalogs in the Computer Age. American Association for State and Local History, Nashville.

Cutbill, John L.

1974 Untitled manuscript being prepared for publication.

Gaines, Sylvia W.

1974 Computer use at an archaeological field location. American Antiquity 39: 454-462.

Gardin, Jean-Claude, and Jacques Le Maître

1974 Le System Documentaire "SOFIA". Centre de Recherches Archéologiques, Marseilles.

Greenberger, Martin, Julius Aronofsky, James L. McKenney, and William F. Massey

1973 Computer and Information Networks. Science 182: 29-35.

1974 Sharing Computer and Information Resources Nationwide. MIT Press.

Heller, Jack

1974 On logical data organization, card catalogs, and the GRIPHOS management information system. Museum Data Bank Research Report 3, Museum Data Bank Committee, Rochester, N.Y.

Scholtz, Sandra

1974 Data structure and computerized museum catalogs. Museum Data Bank Research Report 2, Museum Data Bank Committee, Rochester, N.Y.

Scholtz, Sandra and Robert G. Chenhall

n.d. Archaeological Data Banks in Theory and Practice. (accepted for publication). American Antiquity.

Vance, David

1970 Museum Data Banks. Information Storage and Retrieval 5: 203-211.

1973 Computers in the Museum. IBM data Processing Application GE20-0406-0.

1974 Manual for MCN data preparation. Computer prepared ms. dated 1 June 1974.